KODNING AV MAXIMALA GRAMMATISKA ENHETER

Manual

Jens Allwood Maria Björnberg Alexandra Weilenmann

Version 1, 19 april 1999

---

1. Principer för kodning av maximala grammatiska enheter

När man kodar maximala grammatiska enheter så ska man koda enligt principen att först och främst försöka hitta så stora enheter som möjligt, d.v.s i första hand hela satser. Detta kodas i schemat ”sentences.pl”. I talspråk förekommer många uttryck som inte är satser, för dessa skall man i andra hand försöka hitta fullständiga fraser, och koda det under schemat ”phrases.pl”. I tredje hand ska man, om det inte går att hitta satser eller fraser, koda varje enskilt ord med relevant ordklass. Detta görs i schemat ”parts_of_speech.pl”. Vart och ett utav dessa scheman innehåller olika kategorier och denna manual är en genomgång av vad som bör kodas under respektive kategori.

2. Kodningsschemat sentences.pl

Kodningsschemat sentences.pl innehåller följande kategorier:

declarative_s/1
disj_question/1
exclamative_s/1
imperative_s/1
wh_question/1
yes/no_question/1

Alla fullständiga satser kodas under detta schema. Om satsen innehåller pauser, tvekljud eller upprepningar och liknande, så ska man bortse från dessa och ändå koda den som en sats. OCM-fenomen markeras också för sig och kodas ovanpå satsen, se vidare avsnitt 3. Indirekt tal räknas också som tillhörande satsen. Så räknas t.ex. ”Han sa att han kommer” som en sats.

Deklarativa satser kodas som declarative_s. Tre olika typer av frågor kan man koda: disj_question (disjunktiva frågor), wh_question (frågor som börjar med frågepronomen eller
-adverb) och yes/no-question (frågor som man kan svara ja eller nej på). Man kan också koda exklamativa och imperativa satser.

3. Kodningsschemat phrases.pl

3.1 Kodningsschema

Kodningsschemat phrases.pl innehåller följande kategorier:

adjp/1
advp/1
conj/1
fbp/1
np/1
nump/1
pp/1
subordinate_clause/1
vp/1

3.2 Adjp

Adjp står för adjective phrase och här kodas alltså adjektivfraser.

3.3 Advp

Advp står för adverb phrase. Hit räknas adverbfraser men också adverbial av olika slag, såvida de inte kan räknas som prepositionsfraser. Ett exempel:

$P: ja:0 1<typ den: ä1: tredje fjärde april>1 / 2<nån gång där>2

Här kodas både <1> och <2> som advp.

3.4 Conj

Under conj kodas konjunktionsfraser, t.ex. ”för att”, ”så att”.

3.5 Fbp

Fbp står för feed-back phrase (återkopplingsfras). För att något ska betraktas och kodas som en återkopplingsfras i MaxGram, gäller att frasen måste innehålla ett primärt återkopplingsord. Till primära återkopplingsord räknas:

ja
jo
nej
nä
nja

m
okej
va

Detta innebär alltså att ”just det” och ”just precis” inte ska kodas som återkopplingsfraser. Dessa två fall kodas i stället som adverbfraser. När ett feedback-ord kombineras med andra ord, som i ”ja just det jo” eller ”ja visst” så kan allt kodas som en fbp, beroende på om det verkar ha sagts sammanhängande eller inte. Om det t.ex. är en paus mitt i, så ska orden kodas för sig istället.

3.6 Np

Under np kodas nominalfraser. Observera att detta gäller nominalfraser som består av flera ord! Ett pronomen eller ett egennamn kan i sig utgöra en nominalfras men dessa ska kodas som pronomen respektive nomen under parts_of_speech-schemat. Komplexa nominalfraser som är adverbial kodas som advp.

3.7 Nump

Under nump kodas räkneordsfraser, t.ex. telefonnummer eller andra uppräkningar av tal.

3.8 Pp

Pp står för preposition phrase och här kodas prepositionsfraser. Även om en prepositionsfras utgör adverbial så ska det kodas som prepositionsfras.

3.9 Vp

Under vp kodas fullständiga verbfraser, alltså verb som är ihopsatt med andra ord, eller flera verb. Enstaka intransitiva verb skall kodas som v under parts_of_speech-schemat.

3.10 Subordinate Clause

Under subordinate_clause skall hela bisatser kodas.

När vi stöter på yttranden eller delar av yttranden som inleds med en konjunktion eller subjunktion, ska vi titta på om det är frågan om en samordning med ett tidigare yttrande eller en underordning. I det fall där det rör sig om en samordning, ska konjunktionen kodas för sig under parts_of_speech - conj och satserna för sig som hela satser. Detsamma gäller samordning av Vp eller Np. När det gäller underordning, bör bisatser med subjunktion (t.ex. ”att”, ”om”) föras ihop till en större helhet och kodas som ”subordinate_clause” under ”phrases”.

4. Kodningsschemat parts_of_speech.pl

4.1 Kodningsschemat

Kodningsschemat parts_of_speech.pl innehåller följande kategorier:

adj/1 - adjektiv
adv/1 - adverb
art/1 - artiklar
conj/1 - konjunktioner och subjunktioner
fb/1 - feedback-ord, se 3.5
inf/1 - infinitivmärke
interj/1 - interjektioner
n/1 - nomen
num/1 - räkneord
ocm/1 - OCM-ord
part/1 - partiklar
prep/1 - prepositioner
pron/1 - pronomen
v/1 - verb

4.2 Kodning av OCM i MaxGram

4.2.1 Vad kodas som OCM i MaxGram?

I OCM-manualen finns exempel på många olika OCM-fenomen som förekommer i olika strukturer, t.ex. upprepningar, tvekljud och pauser. När vi kodar MaxGram är det bara speciella ord som alltid eller ofta har OCM-funktion, som ska kodas som OCM. Detta är t.ex. ”e1”,”m1” och andra tvekljud. ”Liksom” har ofta OCM-funktion. Dessa ord kodas under parts_of_speech -schemat och markeras som OCM.

OBSERVERA att feedback-ord ofta har OCM-funktion, men i MaxGram ska dessa alltid kodas som feedback (fb), oavsett funktion (se 2.3.1).
Pauser ska inte heller kodas som OCM, utan kodas bara som ingående i större strukturer.

4.2.2 Kodning av OCM inne i satser

När någonting som i övrigt har strukturen av en fullständig sats innehåller OCM-uttryck, ska vi koda satsen som en sats och därefter markera OCM inuti satsen. Vi ska alltså inte dela upp en sats i mindre delar därför att någon stakar sig eller upprepar sig mitt i satsen.
Ett exempel på detta:
<1 ja1 tänker mej nån som skrive <2 e1>2 dikter om0 naturen>1
där <1> är en deklarativ sats och <2> OCM.

4.2.3 Kodning av OCM inne i fraser

Samma regel som gäller för kodning av OCM inne i satser, gäller även för OCM-fenomen inne i fraser. Vi ska alltså koda OCM även i fraser. Det kan se ut så här:
$R: <1 en riktig <2e1>2 slutsats>1
där <1> är en np och <2> är ett OCM-ord.

4.2.4 Kodning av OCM initialt i fraser

Om ett OCM-fenomen förekommer initialt i en fras, ska vi koda det såsom tillhörande frasen.
Ex. A: <1<2 e1>2 lite grann åt de0 där hållet igen>1
där <1> är en advp och <2> OCM.

4.3 Kodning av återkoppling i MaxGram

4.3.1 Fb-ord som inte har återkopplingsfunktion

När vi kodar MaxGram, ska alla återkopplingsord kodas som återkoppling, oavsett om de förefaller ha en annan funktion, t.ex. OCM. Däremot ska de inte kodas ovanpå satser, som man gör med OCM. Om ett feedback-ord ingår i en sats så kodas endast satsen som helhet.

4.3.2 Yttranden bestående av feedback plus en sats

I yttranden som t.ex. ”ja0 de0 vill ja1” så ska man koda ”ja0” för sig som fb och ”de0 vill ja1” för sig som en deklarativ sats. Detta gäller även om feedbacken står i slutet av yttrandet. Observera att denna feedback också kan vara feedback-elicitorer som ”eller”, ”väl” och ”eller hur”.

5 Övriga principer för MaxGram-kodning

5.1 Direkt tal

Direkt tal kan ses som del i en sats, som i exemplet nedan.

$A: ja1 kan tänka mej att en0 urneurotisk person säjer / å4 ja1 har ju en0 naturli känsla av att de0 e0 si å0 så0 va0

är således en deklarativ sats.

5.2 Yttranden bestående av en sats plus ett namn eller nomen

När det gäller yttranden bestående av ett namn plus en sats, så ska man skilja på om namnet är ett tilltal eller inte. Om det är frågan om tilltal så kodas det inte som en del av satsen. I t.ex. yttrandet ”va1 tror du om den analysen Jonas” så ska ”va1 tror du om den analysen” kodas som en fråga (wh-question) och ”Jonas” som ett nomen, utanför frågan. Om man däremot säger ”Astrid Lindgren hon e0 en bra författare” så ska allting kodas som en deklarativ sats. Detta gäller även i satser med pronomen, som t.ex. ”de0 va1 de0 de0”.
I andra fall när ett substantiv står innan en i sig fullständig sats, ska man se substantivet såsom del i den efterföljande satsen. Ett exempel:

$G: respekt de0 hänger no också mer ihop me0 auktoriteter

Här kodas hela yttrandet som en deklarativ sats. Man ska alltså inte koda substantivet för sig, och resten som en sats.

5.3 När ett ord är utelämnat

Ibland så är ett ord inte transkriberat fastän det ändå har sagts – det har ”flutit ihop” med ett annat ord. Ett exempel här är ”de0 kallt ida”. ”de0” står för ”det är” och ska också kodas som det. Alltså blir ovanstående yttrande en deklarativ sats. Detta gäller inte när som helst när ett ord hoppats över, utan bara när det ”flutit ihop” med något annat, d.v.s det låter inte som om det fattas något ord.

5.4 Utländska ord

Alla utländska ord ska kodas med den ordklass de skulle fått om de hade varit svenska ord.

5.5 Avbrutna ord

Om det går att gissa vilket ord det är som avbrutits, så ska det kodas som om det vore det ordet. Om det inte går att gissa, så ska det kodas som OCM.

5.6 När transkriptören har gjort fel

Om man upptäcker att transkriptören uppenbarligen transkriberat fel, ska man göra en anteckning om det och koda som det grammatiskt står så länge, alltså **inte** som man tror att det ska vara! För att rätta dessa felen måste man gå igenom en speciell procedur och därför får man inte försöka sig på det själv, utan anteckningarna om felen ska lämnas till någon som har lärt sig hur man ändrar dem. Fråga den som är ansvarig för kodningen.

5.7 Ord inom parentes

Ord inom parentes, d.v.s som transkriptören inte varit säker på, skall kodas som om parentesen inte varit där.